紫花苜蓿叶绿体基因组密码子偏好性分析

喻凤*，韩明

（贵州师范大学 生命科学学院，贵阳 550025）

摘要：为了理解紫花苜蓿叶绿体基因组密码子偏好性的使用模式，该文以紫花苜蓿叶绿体基因组中筛选到的 49 条蛋白质编码序列为研究对象，利用 CodonW、CUSP、CHIPS、SPSS 等软件对其密码子的使用模式和偏好性进行分析。结果表明：紫花苜蓿叶绿体基因的第 3 位密码子的平均 GC 含量为 26.87%，有效密码子数（ENC）在 40.6～51.41 之间，多数密码子的偏好性较弱。相对同义密码子使用度（RSCU）分析发现，RSCU＞1 的密码子数目有 30 个，以 A、U 结尾的有 29 个，说明了紫花苜蓿叶绿体基因组 A 或 U 出现的频率较高。中性分析发现，GC_3 与 GC_{12} 的相关性不显著，表明密码子偏性主要受到自然选择的影响；而 ENC-plot 分析发现一部分基因落在曲线的下方及周围，表明突变也影响了部分密码子偏性的形成。此外，有 17 个密码子被鉴定为紫花苜蓿叶绿体基因组的最优密码子。紫花苜蓿叶绿体基因组的密码子使用的偏好性可能受到自然选择和突变的共同作用。该研究将为紫花苜蓿叶绿体基因工程的开展和目标性状的遗传改良奠定基础。

关键词：紫花苜蓿，叶绿体基因组，密码子偏好性，最优密码子

中图分类号：Q943.2　　　　**文献标识码：**A

Analysis of codon usage bias in the chloroplast genome of alfalfa (*Medicago sativa*)

YU Feng*, HAN Ming

（*School of Life Sciences*, *Guizhou Normal University*, Guiyang 550025, China）

Abstract: To character the codon usage patterns of chloroplast genome in alfalfa, forty-nine coding DNA sequences were analyzed using CodonW, CUSP, CHIPS and SPSS software. The results showed that the average GC content at the 3rd codon position was 26.87% in the alfalfa chloroplast genome. The effective codon number (ENC) ranged from 40.6 to 51.41 indicates that codon bias was weak in the alfalfa chloroplast genome. Thirty codons showed relative synonymous codon usage (RSCU) value greater than 1, including 29 AT ending codons, which suggested that the third codon position preferred A or U. Neutrality analysis showed no significant correlation was observed between GC_3 and GC_{12}, suggesting natural selection mainly contributed to the codon usage bias. However, ENC-plot analysis showed that there are some genes with low ENC values lying below or near the expected curve. It indicated that mutation also affected the

---

收稿日期：2020-08-11

基金项目：贵州师范大学 2017 年博士科研启动项目（GZNUD[2017]16）[Supported by Doctoral Foundation of Guizhou Normal University (GZNUD[2017]16)]。

作者简介：喻凤(1988 -)，女，重庆巴南区人，博士，讲师，研究方向为植物细胞遗传学，（E-mail） yufengcaocao@163.com。

***通信作者**

formation of codon usage bias. In addition, 17 codons were identified as optimal codons in alfalfa. We concluded that the codon usage bias of alfalfa is formed under effect of natural selection and mutational bias. This study will lay a good foundation for development of chloroplast genetic engineering and molecular breeding of alfalfa.

Key words: *Medicago sativa*, chloroplast genome, codon usage bias, optimal codons

紫花苜蓿（*Medicago sativa*）是一种多年生豆科牧草，由于其具有适口性好，蛋白质含量高等优点，享有“牧草之王”的美誉（杨青川和孙彦，2011）。其栽培范围广布世界各地，是一种备受关注的世界性优质牧草。研究表明，以紫花苜蓿作为饲料饲养的畜禽，可以提高畜禽的生产、繁殖，肉品质以及机体的免疫力（李刚，2019）。近年来，随着我国畜牧业的发展，对优质牧草的需求量也日益增加。但我国紫花苜蓿等牧草产业的发展相对滞后，产量不足以满足我国市场需求，仍需进口大量牧草。这使得紫花苜蓿在我国的种植面积不断增加，然而，种植过程中田间病虫害和杂草等问题却时有发生，严重时会导致大幅减产和品质下降，制约着我国苜蓿产业的发展（边佳辉等，2020）。传统的育种方法在解决这些难题时收效甚微，而基因工程育种技术的应用却成功获得了大量具有抗病虫和抗除草剂等优良品质的作物材料（杨朝国等，2010；郭三堆等，2015）。由于紫花苜蓿为四倍体植物，遗传背景复杂，与利用核基因工程相比，叶绿体基因工程则具有外源基因可以定点整合、基因拷贝数多、母系遗传等优点（程霞英，2004；周菲等，2015；Jin & Daniell, 2015）。因此，利用叶绿体基因工程引入抗病、抗虫、抗除草剂等基因到紫花苜蓿中对其进行对遗传性状改良，将对畜牧业的可持续发展具有十分重要的意义。

在叶绿体基因工程育种中，有一个关键的环节就是外源基因的表达。而研究表明外源基因的表达水平会受到物种密码子偏好性的影响（Hershberg & Petrov, 2008; Plotkin & Kudla, 2011；Quax et al., 2015）。对某一种生物体的密码子偏好性使用模式进行研究，再根据该生物的最优密码子而设计外源基因表达载体，可以起到优化外源基因在叶绿体基因组中的表达的作用。因此，本文将分析紫花苜蓿叶绿体基因组密码子偏好性的特征及其密码子偏性的影响因素，并确定其最优密码子，为紫花苜蓿叶绿体基因工程的开展和目标性状的遗传改良奠定基础。

1 材料与方法

1.1 叶绿体基因组序列的获得

完整的紫花苜蓿叶绿体基因组下载于 NCBI 数据库（GenBank 登录号为 KU321683.1）（Tao et al., 2017），从该叶绿体基因组中共获得 75 条编码蛋白质的基因序列。为了更加准确分析密码子偏好性，首先剔除小于 300 bp 的基因编码序列，其次选择起始密码子为 ATG，终止密码子为 TAA、TAG、TGA 的基因编码序列，最终一共获得 49 条序列用于后续的数据分析（杨国锋等，2015；刘慧等，2017）。

1.2 分析方法

1.2.1 相对同义密码子使用度分析

使用 CodonW 软件对筛选出的 49 条序列的相对同义密码子使用度（relative synonymous codon usage, RSCU）进行计算。当 RSCU 值等于 1，表示该密码子使用无偏好性；RSCU 值大于 1，表示该密码子的使用频率比其他同义密码子高，反之亦然（Sharp & Li, 1987）。

1.2.2 中性绘图分析

为了分析有关密码子偏好性的影响因素，利用 MEGA6 软件分别统计密码子 3 个位置的 GC 含量，GC_1、GC_2 和 GC_3 分别表示为第一、第二和第三位密码子的 GC 含量。GC_{3S} 表示同义密码子第三位的 GC 含量，GC_{12} 则表示第一位和第二位密码子的 GC 含量的平均值。中

性绘图分析时，以 GC_3 为横坐标、GC_{12} 为纵坐标进行绘图。

1.2.3 ENC-plot 绘图分析

为了进一步分析碱基组成对密码子偏好性的影响，ENC-plot 绘图分析以 GC_{3S} 为横坐标、有效密码子数（effective number of codon，ENC）为纵坐标进行绘制散点图。绘图中的 ENC 标准曲线计算公式为 $ENC = 2+GC_{3S}+29/[GC_{3S}^2+(1-GC_{3S})^2]$（Wright, 1990）。

1.2.4 奇偶偏好（parity rule 2, PR2）分析

为了评估密码子第三位碱基 A 与 T 或 C 与 G 之间的组成偏差是否对密码子偏好性产生影响，对具有 4 个同义密码子编码的氨基酸的各密码子第三位的碱基组成进行作图分析。计算每个基因 $A_3/(A_3+T_3)$ 和 $G_3/(G_3+C_3)$后，分别以 $A_3/(A_3+T_3)$ 和 $G_3/(G_3+C_3)$作纵坐标和横坐标作图，平面图中心表示 A=T，同时 G=C（Sueoka, 2001）。

1.2.5 最优密码子的确定

本文以 ENC 值为偏好性标准确定最优密码子。首先，将 ENC 值进行排序，然后从 ENC 值最高和最低的两极各选 10%的基因，分别建成高低偏性库（低 ENC 值对应高偏性库），并分别计算 RSCU 值和高低两库的Δ RSCU。最后，选取两库Δ RSCU ＞ 0.08，且 RSCU>1 的密码子定义为最优密码子（王鹏良等，2019）。

2 结果与分析

2.1 密码子组成分析

将紫花苜蓿叶绿体基因组中的 49 条编码序列作为一个整体进行碱基组成分析，结果表明其平均 GC 含量为 36.55%，GC_1、GC_2、GC_3 含量分别为 45.66%、37.13%和 26.87%。由此可以看出密码子不同位置的碱基组成的 GC 分布并不是均匀的，其中 GC_1 含量最高，而 GC_3 含量最低。

有效密码子数测量的是某基因的密码子偏好程度，理论上的取值范围为 21~60。当某一个基因平均地使用每一个密码子，则其有效密码子数为 60；当某一个基因只使用每组同义密码子中的一个，则其有效密码子数为 20。参考前人研究以 ENC 值为 35 作为偏性强弱的区分标准（Comeron & Aguadé, 1988）。本研究结果显示，紫花苜蓿叶绿体基因组 ENC 取值范围都在 40.6 以上，因此紫花苜蓿叶绿体基因密码子偏性普遍较弱（表 1）。

表 1 紫花苜蓿叶绿体基因组密码子不同位置的 GC 含量

Table 1 GC contents of different positions on chloroplast genome codon of *Medicago sativa*

基因 Gene	GC 含量 GC content（%）					ENC
	GC_{all}	GC_1	GC_2	GC_3	GC_{3S}	
*clp*P	38.95	55.61	34.18	27.04	22.00	50.13
*psb*B	42.76	55.21	45.97	27.11	23.10	45.28
*pet*B	40.59	47.69	41.67	32.41	26.60	44.97
*pet*D	39.54	51.55	37.89	29.19	26.10	46.57
*rpo*A	31.94	43.71	29.94	22.16	20.60	43.20
*rps*11	42.45	52.52	53.96	20.86	18.70	45.49
*rpl*14	35.23	47.15	38.21	20.33	19.20	45.65
*rpl*16	43.38	51.47	50.74	27.94	23.00	42.27
*rps*3	34.26	44.65	34.88	23.26	21.20	45.63
*rpl*2	42.15	51.45	47.46	27.54	25.70	51.24
*ycf*2	35.42	40.67	32.14	33.46	31.10	51.41
*ndh*B	36.17	41.58	38.54	28.40	24.40	47.51
*rps*7	40.60	51.28	44.87	25.64	22.30	43.17

*rps*12	42.20	52.42	48.39	25.81	25.40	43.60
*ycf*1	30.63	37.28	27.73	26.87	24.60	48.58
*ndh*H	36.55	51.02	35.28	23.35	17.80	47.73
*ndh*A	32.97	41.03	38.04	19.84	17.20	42.02
*ndh*I	33.54	42.59	35.19	22.84	19.50	43.93
*ndh*G	32.58	41.81	33.33	22.60	20.00	45.04
*ndh*E	33.99	42.16	33.33	26.47	22.70	50.30
*ccs*A	31.79	33.95	35.19	26.23	21.20	48.05
*ndh*F	30.87	36.11	34.63	21.88	18.30	43.61
*psb*A	41.62	50.00	42.94	31.92	27.50	41.48
*mat*K	30.44	36.09	30.37	24.85	23.10	48.23
*rbc*L	42.51	57.35	43.70	26.47	23.50	44.92
*atp*B	42.23	57.43	41.75	27.49	25.60	46.99
*atp*E	38.65	48.55	37.68	29.71	26.50	50.67
*ndh*C	33.88	44.63	31.40	25.62	19.80	41.75
*ndh*K	36.99	42.54	42.11	26.32	23.70	51.36
*ndh*J	37.95	49.69	36.48	27.67	23.50	44.12
*rps*4	37.13	51.49	38.12	21.78	20.30	43.55
*ycf*3	38.86	46.75	38.46	31.36	28.60	50.82
*psa*A	41.37	51.91	42.69	29.51	25.30	49.29
*psa*B	40.45	48.71	42.99	29.66	25.00	47.82
*rps*14	40.59	45.54	46.53	29.70	27.10	45.42
psbC	42.48	52.95	46.20	28.27	24.30	43.70
*psb*D	42.66	52.54	43.22	32.20	27.80	45.63
*rpo*B	37.85	49.58	37.82	26.14	23.90	47.77
*rpo*C1	36.28	48.48	37.59	22.79	20.20	45.86
*rpo*C2	34.26	43.11	35.45	24.22	22.20	46.50
*rps*2	36.99	43.88	42.19	24.89	21.00	45.67
*atp*I	35.35	46.77	35.08	24.19	21.80	43.65
*atp*F	36.10	45.08	33.16	30.05	28.70	51.15
*atp*A	39.33	54.79	39.73	23.48	22.00	44.61
*acc*D	35.11	43.91	35.73	25.71	22.90	44.17
*cem*A	31.88	35.65	28.70	31.30	27.60	46.77
*pet*A	38.21	52.34	36.14	26.17	24.80	49.07
*rps*18	30.93	33.33	36.04	23.42	21.30	40.60
*rpl*20	34.27	35.92	37.32	29.58	26.50	49.86
平均值 Average	37.20	46.57	38.59	26.44	23.37	46.26

注: GC_1、GC_2、GC_3 分别表示密码子第一、第二和第三位碱基 GC 含量；GC_{3S} 表示同义密码子第三位碱基 GC 含量；GC_{all} 表示密码子总 GC 含量。

Notes: GC_1, GC_2, GC_3 mean GC value of the first, second, third position of codons, respectively; GC_{3S} means GC value of third position of synonymous codons; GC_{all} means GC value of each gene.

碱基组成和有效密码子数的相关性分析结果显示：GC_{all} 和 GC_1、GC_2、GC_3 的相关性都达到了极显著水平，并且第一位碱基组成和第二位碱基组成的相关性也达到了极显著水平，

但两者与第三位碱基组成的相关性均未达到显著水平（表 2）。表明紫花苜蓿叶绿体编码基因密码子第一位和第二位碱基组成较相似，但与第三位碱基组成却存在差异；而有效密码子数仅与 GC_3 的相关性达到极显著水平，说明密码子第三位上的碱基组成对紫花苜蓿叶绿体编码基因的密码子偏好性影响显著。

表 2 紫花苜蓿叶绿体基因组中相关参数的相关性分析

Table 2 Correlation analysis of related parameters on chloroplast genome of *Medicago sativa*

	GC_1	GC_2	GC_3	GC_{all}	ENC
GC_2	0.599**		0.106	0.848**	-0.177
GC_3	0.097	0.106		0.394**	0.410**
GC_{all}	0.865**	0.848**	0.394**		0.002
ENC	-0.550	-0.177	0.410**	0.002	

注：** 表示在 0.01 水平上显著相关。

Note: ** means significant differences at the 0.01 level.

2.2 中性绘图分析

中性绘图分析结果如图 1 所示，由图可知 GC_{12} 分布于 0.321 8～0.532 4，GC_3 分布于 0.198 4～0.334 6，几乎各基因的坐标点都未落在对角线上或沿着对角线分布。GC_{12} 和 GC_3 的相关系数为 0.113 3，统计结果表明两者的相关性不显著。回归曲线斜率为 0.181 2，表明 GC_{12} 和 GC_3 的相关性较弱。进一步说明突变对密码子偏性的形成的影响较弱，其他因素（如选择）可能对密码子的偏好性形成影响较大。

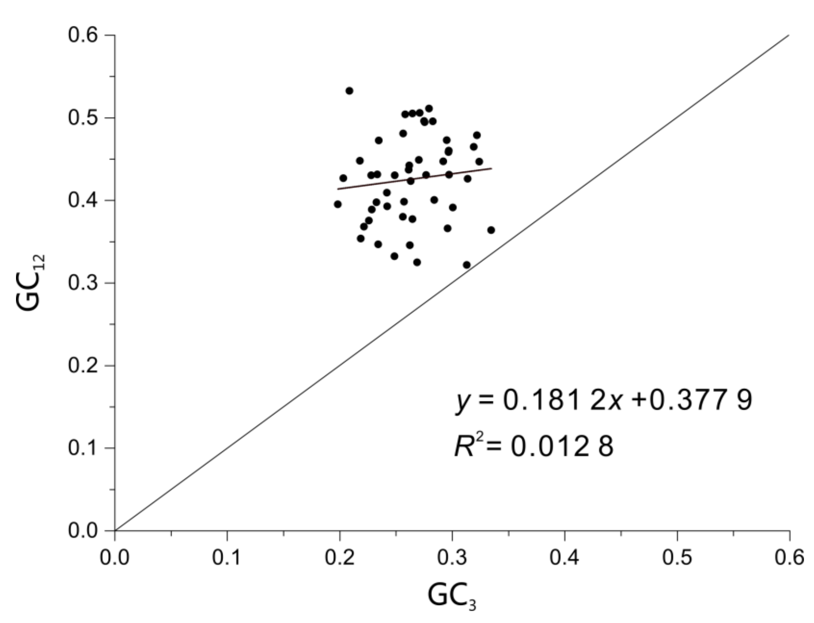


图 1 紫花苜蓿叶绿体基因中性绘图分析

Fig. 1 Neutrality plot analysis of chloroplast genes of *Medicago sativa*

2.3 ENC-plot 绘图分析

为了进一步分析碱基组成对密码子偏好性的影响，ENC-plot 绘图以有效密码子数为纵坐标，GC_{3S} 为横坐标，定位所有基因的坐标系，同时添加标准曲线。结果显示大部分基因沿标准曲线分布或位于标准曲线附近（图 2）。为了进一步反映差异，根据 ENCexp 公式求出 ENC 的理论值，再根据公式 $ENC_{ratio}=(ENC_{exp}-ENC_{obs})/ENC_{exp}$ 求算出 ENC 比值。在此基础上分析所有参与研究基因的 ENC 频数分布（表 3）。统计结果表明，大多数的基因的 ENC 比值分布在-0.05~0.15 之间，该范围内实际 ENC 与期望 ENC 值相差较小。因此，说明突变对紫花苜蓿叶绿体基因组的部分基因密码子偏好性的形成影响较大。

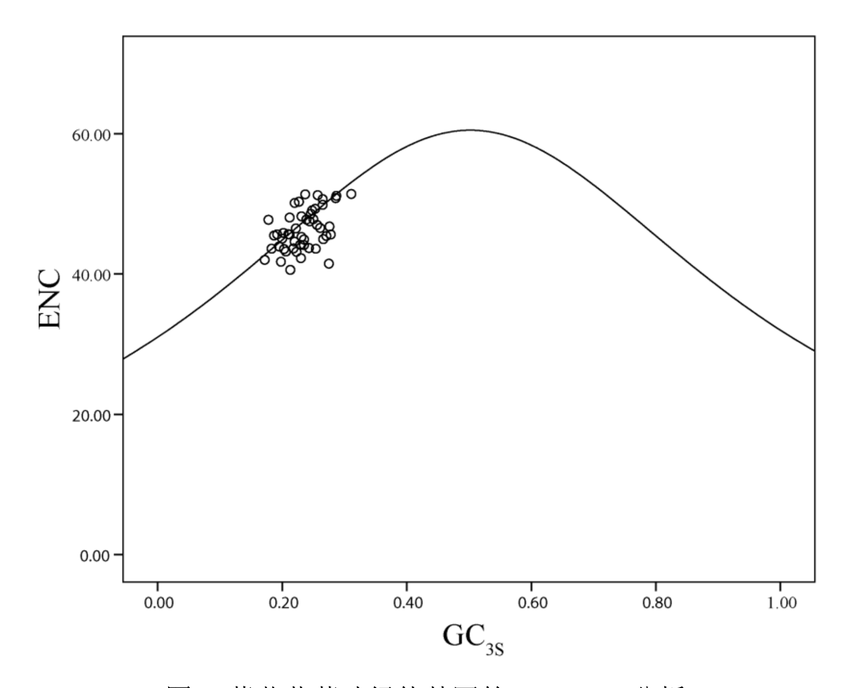


图 2 紫花苜蓿叶绿体基因的 ENC-plot 分析

Fig. 2 ENC-plot analysis of chloroplast genes of *Medicago sativa*

表 3 ENC 比值频数分布

Table 3 Distribution of ENC ratio

组限 Class boundary	组中值 Class middle value	组数 Number	组频 Frequency
-0.15 ~ -0.05	-0.1	4	0.081 6
-0.05 ~ 0.05	0	29	0.591 8
0.05 ~ 0.15	0.1	15	0.306 1
0.15 ~ 0.25	0.2	1	0.020 4

2.4 PR2-plot 分析

利用 PR2-plot 绘图分析了紫花苜蓿叶绿体基因组中部分氨基酸家族中密码子第三位碱基 A 与 T、C 与 G 之间的关系（图 3）。结果显示，大部分基因分布在图的右下方，表明密码子第三位 T 的使用频率高于 A，G 的使用频率高于 C。如果密码子使用模式仅受到突变的影响，则 A 和 T、G 和 C 的碱基使用频率理论上应该相等。但本研究的结果显示 A 和 T、G 和 C 的使用存在不均衡，表明紫花苜蓿叶绿体密码子的使用模式不仅只受到突变的影响，还可能受到其他因素，如选择等的影响。

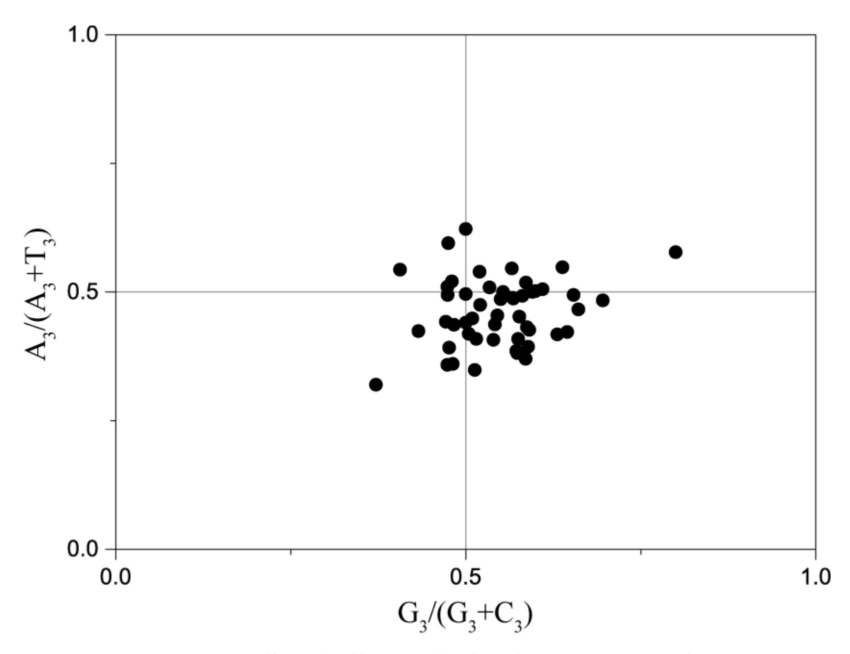


图 3 紫花苜蓿叶绿体基因的 PR2-plot 分析

Fig. 3 PR2-plot analysis of chloroplast genes of *Medicago sativa*

2.5 相对同义密码子使用度分析

紫花苜蓿叶绿体基因组中各氨基酸的相对同义密码子的使用度分析如表 4 所示，结果显示共有 30 个密码子的 RSCU 大于 1。其中以 U 或 A 结尾的密码子占比为 96.67%，以 U 结尾的为 13 个，以 A 结尾的为 16 个，说明紫花苜蓿偏爱使用结尾为 U 和 A 的密码子。

表 4 紫花苜蓿叶绿体基因组中各氨基酸的 RSCU 分析

Table 4 RSCU analysis of protein coding region in chloroplast genome of *Medicago sativa*

氨基酸 Amino acid	密码子 Codon	数目 Codon	RSCU
苯丙氨酸 Phe	UUU	814	1.42
	UUC	330	0.58
丝氨酸 Ser	UCU	436	1.77
	UCC	217	0.88
	UCA	287	1.17
	UCG	135	0.55
酪氨酸 Tyr	UAU	629	1.66
	UAC	128	0.34
色氨酸 Trp	UGG	346	1.00
半胱氨酸 Cys	UGU	157	1.50
	UGC	52	0.50
亮氨酸 Leu	UUA	714	2.05
	UUG	432	1.24
	CUU	436	1.25
	CUC	100	0.29
	CUA	289	0.83
	CUG	119	0.34
脯氨酸 Pro	CCU	341	1.68
	CCC	141	0.69

	CCA	241	1.18
	CCG	91	0.45
组氨酸 His	CAU	363	1.55
	CAC	104	0.45
谷氨酰胺 Gln	CAA	574	1.61
	CAG	140	0.39
精氨酸 Arg	CGU	271	1.44
	CGC	81	0.43
	CGA	254	1.35
	CGG	71	0.38
异亮氨酸 Ile	AUU	901	1.53
	AUC	297	0.51
	AUA	565	0.96
苏氨酸 Thr	ACU	421	1.70
	ACC	164	0.66
	ACA	308	1.25
	ACG	96	0.39
天冬酰胺 Asn	AAU	784	1.58
	AAC	208	0.42
赖氨酸 Lys	AAA	876	1.59
	AAG	229	0.41
丝氨酸 Ser	AGU	325	1.32
	AGC	74	0.30
精氨酸 Arg	AGA	331	1.75
	AGG	125	0.66
缬氨酸 Val	GUU	413	1.51
	GUC	113	0.41
	GUA	426	1.56
	GUG	141	0.52
丙氨酸 Ala	GCU	517	1.88
	GCC	147	0.54
	GCA	314	1.14
	GCG	121	0.44
天冬氨酸 Asp	GAU	699	1.61
	GAC	167	0.39
谷氨酸 Glu	GAA	858	1.55
	GAG	247	0.45
甘氨酸 Gly	GGU	473	1.40
	GGC	119	0.35
	GGA	559	1.66
	GGG	196	0.58
甲硫氨酸 Met	AUG	432	1.00
终止密码子 TER	UAA	30	1.84
	UAG	8	0.49

UGA	11	0.67

2.6 最优密码子的确定

本文根据Δ RSCU 法确定最优密码子。首先，将 ENC 值进行排序，然后从 ENC 值最高和最低的两极各选 10%的基因，分别建成高低偏性库，并分别计算 RSCU 值和高低两库的Δ RSCU（表 5）。结果发现两库Δ RSCU > 0.08 的一共有 27 个高表达密码子（表 5 中*标明）。选取表 5 中的高表达的密码子和表 4 中的高频密码子相结合进行分析，将其共有密码子定义为最优密码子。最终确定了紫花苜蓿叶绿体基因组的最优密码子 17 个（UGU、UUA、CCU、CCA、CAA、CGU、ACU、ACA、AAA、AGA、GUU、GUA、GCU、GAU、GAA、GGU 和 UAA），在这 17 个最优密码子中，全以 A 或 U 结尾，此结果与 GC_{all} 和 GC_3 的高 A、T 含量一致。

表 5 紫花苜蓿叶绿体基因组最优密码子分析

Table 5 Putative optimal codons in chloroplast genome of *Medicago sativa*

氨基酸 Amino Acid	密码子 Codon	高表达基因 High expressed genes		低表达基因 Low expressed genes		△RSCU
		数目 Number	RSCU	数目 Number	RSCU	
苯丙氨酸 Phe	UUU	48	1.28	101	1.20	0.08
	UUC	27	0.72	67	0.80	-0.08
丝氨酸 Ser	UCU*	37	2.24	68	1.75	0.49
	UCC	11	0.67	35	0.90	-0.23
	UCA	19	1.15	54	1.39	-0.24
	UCG	9	0.55	21	0.54	0.01
酪氨酸 Tyr	UAU	30	1.43	99	1.66	-0.23
	UAC*	12	0.57	20	0.34	0.23
色氨酸 Trp	UGG*	26	2	40	1.90	0.10
半胱氨酸 Cys	UGU*	4	2	21	1.35	0.65
	UGC	0	0	10	0.65	-0.65
亮氨酸 Leu	UUA*	40	2	83	1.68	0.32
	UUG	20	1	65	1.32	-0.32
	CUU	28	1.40	67	1.36	0.04
	CUC	3	0.15	19	0.39	-0.24
	CUA*	26	1.30	41	0.83	0.47
	CUG	3	0.15	21	0.43	-0.28
脯氨酸 Pro	CCU*	24	2.04	43	1.61	0.43
	CCC	6	0.51	21	0.79	-0.28
	CCA*	13	1.11	25	0.93	0.18
	CCG	4	0.34	18	0.67	-0.33
组氨酸 His	CAU	7	1.27	52	1.51	-0.24
	CAC*	4	0.73	17	0.49	0.24
谷氨酰胺 Gln	CAA*	26	1.79	68	1.40	0.39
	CAG	3	0.21	29	0.60	-0.39
精氨酸 Arg	CGU*	18	1.85	26	1.11	0.74
	CGC	6	0.62	13	0.55	0.07
	CGA	13	1.33	41	1.74	-0.41

	CGG	2	0.21	14	0.60	-0.39
异亮氨酸 Ile	AUU	58	1.43	93	1.43	0
	AUC	23	0.57	37	0.57	0
	AUA	30	1.03	74	1.19	-0.16
苏氨酸 Thr	ACU*	17	1.45	41	1.31	0.14
	ACC	10	0.85	26	0.83	0.02
	ACA*	17	1.45	42	1.34	0.11
	ACG	3	0.26	16	0.51	-0.25
天冬酰胺 Asn	AAU	27	1.20	142	1.59	-0.39
	AAC*	18	0.80	37	0.41	0.39
赖氨酸 Lys	AAA*	35	1.79	133	1.37	0.42
	AAG	4	0.21	61	0.63	-0.42
丝氨酸 Ser	AGU	18	1.09	47	1.21	-0.12
	AGC*	5	0.30	8	0.21	0.09
精氨酸 Arg	AGA*	16	2.56	54	2.30	0.26
	AGG	4	0.64	37	1.57	-0.93
缬氨酸 Val	GUU*	30	1.82	44	1.64	0.18
	GUC*	8	0.48	9	0.34	0.14
	GUA*	26	1.58	31	1.16	0.42
	GUG	2	0.12	23	0.86	-0.74
丙氨酸 Ala	GCU*	42	1.47	37	1.47	0.83
	GCC	3	0.79	20	0.79	-0.63
	GCA	21	1.07	27	1.07	0.08
	GCG	7	0.67	17	0.67	-0.29
天冬氨酸 Asp	GAU*	23	1.77	120	1.66	0.11
	GAC	3	0.23	25	0.34	-0.11
谷氨酸 Glu	GAA*	43	1.59	106	1.27	0.32
	GAG	11	0.41	61	0.73	-0.32
甘氨酸 Gly	GGU*	40	1.90	34	0.96	0.94
	GGC	6	0.29	15	0.42	-0.13
	GGA	29	1.38	64	1.80	-0.42
	GGG	9	0.43	29	0.82	-0.39
甲硫氨酸 Met	AUG*	28	0.97	50	0.81	0.16
终止密码子 TER	UAA*	4	0.64	2	0.09	0.55
	UAG*	1	0.16	1	0.04	0.12
	UGA	0	0	2	0.10	-0.10

注：* 表示高表达密码子。

Note: * indicates high expression codons.

3 讨论与结论

在长期进化过程中，特定物种为了适应自身的基因组环境，往往形成了一套具有偏好性的密码子使用模式（李平等，2011）。密码子的使用偏好性受到基因组大小、CpG 岛、GC 分布的位置、基因密度、基因序列长度、基因表达水平、tRNA 丰度、蛋白质二级结构以及密码子变异偏好性等多种因素影响（Hanson & Coller, 2018；Paul et al., 2018）。其中自然选

择和突变是影响密码子使用偏性的两个主要因素，也是目前备受关注的两个因素（Sharp et al., 2010；Rao et al., 2011；石元豹等，2018）。由于密码子第三位碱基突变通常受到的选择压较第一、二位上的碱基小，因此分析密码子第三位碱基的碱基组成对密码子的使用模式研究具有重要意义（原晓龙等，2020）。本研究表明紫花苜蓿叶绿体基因组各基因第三位密码子碱基富含 A 和 T，与蒺藜苜蓿（*Medicago truncatula*）（杨国锋等，2015）、酸枣（*Ziziphus jujuba* var. *spinosa*）（胡晓艳等，2019）和蝴蝶兰（*Phalaenopsis aphrodite* subsp. *formosana*）（续晨等，2010）等植物叶绿体基因组第三位密码子使用偏性相似，表明这类植物的叶绿体基因组密码子使用模式具有高度相似性。这提示我们高等植物叶绿体基因组可能含有较高的 AT 碱基含量，且密码子使用模式的偏好性在进化上存在一定的保守性。

紫花苜蓿的叶绿体基因组编码基因的 ENC 值都大于 40.6，说明紫花苜蓿基因组存在较弱的密码子偏性。中性绘图分析结果表明紫花苜蓿叶绿体基因组密码子 GC_{12} 与 GC_3 相关性不显著，表明密码子的使用偏性可能更多的受到选择的影响；而 ENC-plot 绘图分析表明突变是部分基因密码子偏好性形成的重要因素；PR2-plot 分析的结果则进一步显示密码子的偏好性不仅受到突变的影响还受到选择的影响。因此，在长期的进化过程中，紫花苜蓿叶绿体基因组密码子的使用偏性更多的受到选择的影响，而突变也是部分基因的密码子偏性形成的重要因素。这与前人对剑麻（*Agave hybrid*）（金刚等，2018）、糜子（*Panicum miliaceum*）（刘慧等，2017）和梧桐属植物（陆奇丰等，2020）的叶绿体基因组密码子偏性影响因素分析结果相似，表明影响植物叶绿体基因组密码子偏性的因素是综合性的。但具体的影响机制还有待通过进一步的深入研究来揭示。

本研究将高表达的高频密码子定义为最优密码子，在紫花苜蓿叶绿体基因组中确定的 17 个最优密码子分别为 UGU、UUA、CCU、CCA、CAA、CGU、ACU、ACA、AAA、AGA、GUU、GUA、GCU、GAU、GAA、GGU 和 UAA。这些最优密码子都以 A 或 U 碱基结尾，这与蒺藜苜蓿（*Medicago truncatula*）（杨国锋等，2015）、巨桉（*Eucalyptus grandis*）（王鹏良等，2019）、糜子（*Panicum miliaceum*）（刘慧等，2017）和籽粒苋（*Amaranthus hypochondriacus*）（冯瑞云等，2019）等物种叶绿体基因组最优密码子分析结果一致。同时最优密码子的结果提示在紫花苜蓿的叶绿体基因工程的外源基因载体的设计时，选用这些 A 或 U 碱基结尾最优密码子或可提高外源基因的表达效率。

参考文献：

BIAN JH, LIU ZY, LI ZY, et al., 2020. Advances in alfalfa improvement using genetic engineering[J]. Pratac Sci, 37(1): 139-155. [边佳辉，刘自扬，李宗英，等，2020, 紫花苜蓿基因工程改良研究进展[J]. 草业科学，37（1）：139-155.]

CHENG XY, 2004. Advances in the research of chloroplast genetic transformation[J]. Bull Biol, 39(3): 15-17. [程霞英，2004. 叶绿体遗传转化的研究进展[J]. 生物学通报，39（3）：15-17.]

COMERON JM, MONTSERRAT A, 1998. An evaluation of measures of synonymous codon usage bias[J]. J Mol Evol, 47(3): 268-274.

FENG RY, MEI C, WANG HJ, et al., 2019. Analysis of codon usage in the chloroplast genome of grain amaranth (*Amaranthus hypochondriacus* L.)[J]. Chin J Grassl, 41(4): 8-15. [冯瑞云，梅超，王慧杰，等，2019. 籽粒苋叶绿体基因组密码子偏好性分析[J]. 中国草地学报，41（4）：8-15.]

GUO SD, WANG Y, SUN GQ, et al., 2015. Twenty years of research and application of transgenic cotton in China[J]. Sci Agric Sin, 48(17):3372-3387. [郭三堆，王远，孙国清，等，2015. 中国转基因棉花研发应用二十年[J]. 中国农业科学，48（17）：3372-3387.]

HANSON G, COLLER J, 2018. Codon optimality, bias and usage in translation and mRNA

decay[J]. Nat Rev Mol Cell Biol, 19(1): 20-30.

HERSHBERG R, PETROV DA, 2008. Selection on codon bias[J]. Ann Rev Genet, 42(1): 287-299.

HU XY, YU YQ, HAN YZ, et al., 2019. Codon usage bias analysis of the chloroplast genome of *Ziziphus jujuba* var. *spinosa*[J]. J For Environ, 39(6): 621-628. [胡晓艳，许艳秋，韩有志，等，2019. 酸枣叶绿体基因组密码子使用偏性分析[J]. 森林与环境学报，39（6）：621-628.]

JIN G, TAN X, LONG LY, et al., 2018. Characteristics of codon usage in the chloroplast protein-coding genes of *Agave hybrid* No. 11648[J]. J Fujian Agric For Univ(Nat Sci Ed), 47(6): 705-710. [金刚，谭旭，龙凌云，等，2018. 剑麻叶绿体基因组编码序列密码子的使用特征[J]. 福建农林大学学报(自然科学版)，47（6）：705-710.]

JIN S, DANIELL H. 2015. The engineered chloroplast genome just got smarter[J]. Trends in Plant Sci, 20(10): 622-640.

LI G, 2019. Research progress in *Medicago sativa* L.[J]. Anhui Agric Sci Bull, 25(14): 71-75. [李刚，2019. 紫花苜蓿研究进展[J]. 安徽农学通报，25（14）：71-75.]

LI P, BAI YF, FENG RY, 2011. Analysis of codon bias of NAD-ME gene in *Amaranthus hypochondriacus* [J]. J Appl Environ Biol, 17(1): 12-17. [李平，白云凤，冯瑞云，2011. 籽粒苋苹果酸酶(NAD-ME)基因密码子偏好性分析[J]. 应用与环境生物学报，17（1）：12-17.]

LIU H, WANG MX, YUE WJ, et al., 2017. Analysis of codon usage in the chloroplast genome of broomcorn millet (*Panicum miliaceum* L.)[J]. Plant Sci J, 35(3): 362-371. [刘慧，王梦醒，岳文杰，等，2017. 糜子叶绿体基因组密码子使用偏性的分析[J]. 植物科学学报，35（3）：362-371.]

LU QF, LUO WH, HUANG ZH, 2020. Codon usage bias of chloroplast genome from two species of *Firmiana* Marsili[J]. Guihaia, 40(2): 173-183. [陆奇丰，骆文华，黄至欢，2020. 两种梧桐叶绿体基因组密码子使用偏性分析[J]. 广西植物，40（2）：173-183.]

RAO Y, WU G, WANG Z, et al., 2011. Mutation bias is the driving force of codon usage in the *Gallus gallus* genome[J]. DNA Res, 18(6): 499-512.

PAUL P, MALAKAR AK, CHAKRABORTY S, 2018. Codon usage and amino acid usage influence genes expression level[J]. Genetica, 146(1): 53-63.

PLOTKIN JB, KUDLA G, 2011. Synonymous but not the same: the causes and consequences of codon bias[J]. Nat Rev Genet, 12(1):32-42.

QUAX TE, CLAASSENS NJ, SOELL D, et al., 2015. Codon bias as a means to fine-tune gene expression [J]. Mol Cell, 59 (2); 149 -161.

SHARP PM, EMERY LR, ZENG K, 2010. Forces that influence the evolution of codon bias[J]. Phil Trans Roy Soc London B, 365(1544): 1203-1212.

SHARP PM, LI WH, 1987. The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications[J]. Nucl Acid Res, 15(3): 1281-1295.

SHI YB, WANG GB, YANG XM, et al., 2018. Analysis of codon usage bias of gene factors in *Ginkgo biloba* WRKY family[J]. Mol Plant Breed, 17(5): 1503-1511. [石元豹，汪贵斌，杨晓明，等，2018. 银杏 WRKY 家族基因密码子使用偏向性分析[J]. 分子植物育种，17（5）：1503-1511.]

SUEOKA N, 2001. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses[J]. J Mol Evol, 53(4-5): 469-476.

TAO X, MA L, ZHANG Z, et al., 2017. Characterization of the complete chloroplast genome of

alfalfa (*Medicago sativa*) (Leguminosae)[J]. Gene Reports, 6: 67-73.

WANG PL, WU SC, YANG, LP, et al., 2019. Analysis of codon bias of chloroplast genome in *Eucalyptus grandis*[J]. Guihaia, 39(12): 1583-1592. [王鹏良，吴双成，杨利平，等，2019. 巨桉叶绿体基因组密码子偏好性分析[J]. 广西植物，39（12）：1583-1592.]

WRIGHT F, 1990. The 'effective number of codons' used in a gene[J]. Gene, 87(1): 23-29.

XU C, BEN AL, CAI XN, 2010. Analysis of synonymous codon usage in chloroplast genome of *Phalaenopsis aphrodite* subsp. *formosana*[J]. Mol Plant Breed, 8(5): 945-950. [续晨，贲爱玲，蔡晓宁，2010. 蝴蝶兰叶绿体基因组密码子使用的相关分析[J]. 分子植物育种，8（5）：945-950.]

YANG CG, YANG FF, ZHANG DF, 2010. Methods of screening transgenic maize with herbicide glufosinate resistance[J]. Crops, 269(6):32-35. [杨朝国，杨方方，张登峰，2010. 抗除草剂草丁膦转基因玉米后代筛选方法的研究[J]. 作物杂志，269（6）：32-35.]

YANG GF, SU KL, ZHAO YR, et al., 2015. Analysis of codon usage in the chloroplast genome of *Medicago truncatula*[J]. Acta Pratac Sin, 24(12): 171-179. [杨国锋，苏昆龙，赵怡然，等，2015. 蒺藜苜蓿叶绿体密码子偏好性分析[J]. 草业学报，24（12）：171-179.]

YANG QC, SUN Y, 2011. The history, current situation and development of alfalfa breeding in China[J]. Chin J Grassl, 33(6): 95-101. [杨青川，孙彦，2011. 中国苜蓿育种的历史、现状与发展趋势[J]. 中国草地学报，33（6）：95-101.]

YUAN XL, LI YQ, ZHANG JF, et al., 2019. (2019-09-19) [2020-03-22]. Analysis of codon usage bias in the chloroplast genome of *Dalbergia odorifera*[EB/OL]. http://kns.cnki.net/kcms/detail/45.1134.Q.20190918.1651.008.html. [原晓龙，李云琴，张劲峰，等，2019. 降香黄檀叶绿体基因组密码子偏好性分析[EB/OL]. http://kns.cnki.net/kcms/detail/45.1134.Q.20190918.1651.008.html]

ZHOU F, LU SZ, GAO L, et al., 2015. Plastid genome engineering: novel optimization strategies and applications[J]. Hereditas, 37(8): 777-792. [周菲，路史展，高亮，等，2015. 植物质体基因工程:新的优化策略及应用[J]. 遗传，37（8）：777-792.]